

# 周报

## 本周工作

- 1 本周一小半的时间主要在看论文以及准备组会，这周是我第一次在 vag 小组内来讲稍有紧张，不过据台下的小伙伴说好像讲的不错  $O(\cap \cap)O$ ，总体感觉自己讲的能力有所提升，在台上也能“侃侃而谈”了（台下同学说我讲的太久）
- 2 由于下一周是短学期期末，这一周主要是在看高级操作系统的课堂演讲内容（下周开始操作系统要求组队选题上去讲）
- 3 网易数据方面，本周主要是开始看 spark 并且尝试在目前的 hadoop 平台上搭建 spark。

## 组会总结：

出席：刘致奇 何静轩 刘疏 陈吟吟

内容：

主要总结了一下目前的工作，在雾霾模型方面，考虑游戏是上线或者不上线，如果单纯的用上线不上线表征可能不具代表性，因此不论雾霾与否玩家都极有可能上线，雾霾影响的应该是上线的时间。因此修改基础模型，将原来的是否上线改为上线时间。同时陈吟吟同学说目前利用的统计模型研究雾霾天气对出行影响方面效果不显著，因此让陈吟吟和刘疏同学除了尝试雾霾模型外还加入一些其他的天气成分，并按照加入的数据特点修改模型，例如晴雨天气等。

同时，讨论由于数据量的原因，不同城市可能数据量不同，同时有些省份ip不准确的原因，因此我们讨论先抽取其中比较多的几组数据（具体几所上线人数较多的城市）。来对模型进行分析。

对于数据获得方面，让何静轩同学除了爬取雾霾模型外还顺便将其他天气爬下来。同时为了尽快让其他同学获得数据进行试验，因此先让何静轩同学选取几所城市来尝试。

## 小组本周进度：

**陈吟吟：** 主要在做基站+申请学校。

**刘疏：** 主要是整理了何静轩的数据，筛除了一些干扰性的数据项，从玩家每日上线总时长，平均时长还有上线人数三个数据对象做回归分析，回归对象主要是aqi指数，pm2.5指数还有气温。发现玩家每日上线总时长和aqi指数，pm2.5指数还有气温三个字变量最可能具有线性关系（通过F检验给出），于是我计划根据数据把原始数据分为白天与晚间两个类别，计算各自上线总时长，再和aqi指数，pm2.5指数以及白天、晚间各自的温度做回归分析，应该会有更好的结果。不过工作还没有全部结束，希望有新的结果。

**何静轩：** 本周我的工作主要是在230服务器登录数据中增加了aqi，pm2.5，天气状况，温度等字段。方法是利用爬虫爬取

<http://www.tianqihoubao.com/lishi/beijing/month/201311.html>上的天气数据。如今初步爬取了2013年11月北京市的数据，如果后续有更多数据需要，可以找我。

**沈彦迪：** 利用时序模型对北京市游戏数据与天气模型进行研究，工作尚未完成，同时效果不太显著，因为按照时序模型的来做，最好是有一天的详细天气数据变

化。同时目前城市比较单一，希望下周能有更多的城市。